Available online at www.sciencedirect.com
SCIENCE DIRECT
An International Journal
computers & mathematics with applications
ELSEVIER Computers and Mathematics with Applications 48 (2004) 747–754
www.elsevier.com/locate/camwa

# Search Trees and Stirling Numbers

A. H. MANOB
American International University Bangladesh
Dhaka, Bangladesh

T. M. ISLAM
Department of Computer Science
University of Waterloo
Canada

M. T. PARVEZ AND M. KAYKOBAD
Department of Computer Science and Engineering
Bangladesh University of Engineering and Technology
Dhaka-1000, Bangladesh
kaykobad@cse.buet.ac.bd

*(Received September 2003; revised and accepted March 2004)*

**Abstract**—Search trees, specially binary search trees, are very important data structures that contributed immensely to improved performance of different search algorithms. In this paper, we express certain parameters of search trees in terms of Stirling numbers. We also introduce two new inversion formulas relating Stirling numbers of the first and second kinds. © 2004 Elsevier Ltd. All rights reserved.

**Keywords**—Search trees, Generating functions, Stirling numbers.

## 1. INTRODUCTION

That searching is of paramount importance for computer systems can be recognized from the fact that over 25% of the running time of computers is spent on sorting [1], which is primarily used for searching. The linear complexity of searching in an unsorted array or a linked list, and deleting from or inserting an element into a sorted array has made all of them inappropriate as a data structure in which all of those operations have to be performed. This has resulted in a new data structure called *search tree* in which all those operations can be performed in logarithmic time on the average. Thus, a search tree can be used both as a dictionary and as a priority queue. Although there is a possibility that a search tree can degenerate into a linear list, its average case complexity is still $O(\log n)$. In order to avoid a bad worst case complexity, a modification of search trees, known as *AVL trees*, was introduced by Adel'son-Vel'skii and Landis in 1962 [2]. Several attempts have been made to find out the expected behavior of AVL trees. Foster (see [3, p. 462]) has proposed an approximate model. Yao [4] introduced a different

We profusely thank Professor G.H. Gonnet for *Maple* that relieved us from tedious manipulation of mathematical expressions.


doi:10.1016/j.camwa.2004.03.005

technique called fringe analysis for average case analysis of balanced search trees. Brown [2] later used this technique to obtain a partial analysis of AVL trees. In his analysis, Brown considered the collections of AVL subtrees with three or less leaves and called it the fringe of the AVL tree.

In an earlier paper, [5] we introduced a new recurrence relation capable of capturing a lot more properties of search trees together with their generalization with respect to number of sons. In this paper, we express some of the parameters in terms of Stirling numbers. We also introduce two new inversion formulas in the same spirit as those in Graham [3] relating Stirling numbers of the first and second kinds.

## 2. PRELIMINARIES

We first consider random binary search trees, and assume that each of the $n$ elements is equally likely to be inserted at the root and that for the elements of the left and right subtrees the same is true recursively. In order to deduce values expected height of a binary search tree, the following recurrence relation is available in the literature [3]:

$$E_n = n + 1 + \frac{2}{n} \sum_{i=0}^{n-1} E_i. \tag{2.1}$$

Solution to this recurrence relation yields

$$E_n = 1.38 n \lg(n) + O(n). \tag{2.2}$$

Since binary search trees can degenerate, although with extremely small probability, into a linear list requiring linear time for search, a refinement of binary search tree has been introduced to avoid this pathological case. In this data structure, called *height balanced tree*, additional balancing operation is performed whenever insertion into and deletion from it violate balancing property. Algorithms exist for the implementation of height balanced trees. The bound for the height of the tree is $h \leq 1.44 \lg(n+1)$. Thus, the number of probes will be at most $1.44 \lg(n+1)$ in the worst case.

In the next section, we express some of the parameters of search trees obtained from recurrence relations introduced in [6] in terms of Stirling numbers, and introduce two new inversion formulas relating Stirling numbers of the first and the second kinds.

## 3. SOME NEW RECURRENCE RELATIONS

Let us denote binary and ternary search trees respectively by $T(2)$ and $T(3)$. We also denote $n(n-1)\dots(n-k+1)$ by $n^{\underline{k}}$ following the notation of falling powers, and $n(n+1)\dots(n+k-1)$ by $n^{\overline{k}}$ rising powers as in [3]. The first parameter of our interest is the expected number $E(k,n)$ of external nodes at level $k$ of a search tree on $n$ internal nodes. We will use appropriate superscripts in parentheses to distinguish parameters of binary and ternary search trees. Assuming as mentioned earlier that any element may be inserted in the root with equal probability we have the following relation:

$$E^{(2)}(k,n) = \begin{cases} 1, & \text{if } k = n = 0, \\ \dfrac{2}{n} \displaystyle\sum_{i=0}^{n-1} E^{(2)}(k-1,i), & \text{otherwise.} \end{cases} \tag{3.1}$$

For any empty binary tree there is exactly one external node at level 0, that is at the root. Otherwise, since the search tree consists of a root and two disjoint binary search trees $T_l$ and $T_r$

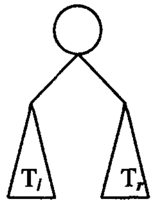


Figure 1. A binary search tree.

as in Figure 1, and since number of elements in $T_l$ and $T_r$ are uniformly distributed in the range from 0 to $n-1$, we have the desired recurrence.

After some manipulations, we obtain

$$nE^{(2)}(k,n) - (n-1)E^{(2)}(k,n-1) = 2E^{(2)}(k-1,n-1). \tag{3.2}$$

Different terms in (3.2) can be found as follows. For example, the first term comes from the coefficient of the following generating function:

$$zG'_z(u,z) = \sum_{k,n\geq 0} nE^{(2)}(k,n)u^k z^n.$$

Now,

$$G(u,z) = \sum_{k,n\geq 0} E(k,n)u^k z^n = \sum_{k,n-1\geq 0} E(k,n-1)u^k z^{n-1}. \tag{3.3}$$

After some manipulations [5], we find that

$$\begin{aligned} zG'_z(u,z) - z^2G'_z(u,z) &= 2uzG(u,z) \\ &\Rightarrow \frac{G'_z(u,z)}{G(u,z)} = \frac{2u}{1-z} \\ &\Rightarrow \ln(G(u,z)) = -2u\ln(1-z) + \ln c = \ln c(1-z)^{-2u} \\ &\Rightarrow G(u,z) = c(1-z)^{-2u}. \end{aligned} \tag{3.4}$$

The coefficient in equation (3.4) is $c$, which equals 1. So, solution to (3.4) becomes

$$\begin{aligned} G(u,z) &= (1-z)^{-2u} \\ &= \sum_n \binom{-2u}{n} z^n \\ &= \sum_n \frac{(-2u)^{\underline{n}}}{n!} z^n \\ &= \sum_n \frac{z^n}{n!} \sum_k \left[{n \atop k}\right] (-2u)^k(-1)^{n-k}. \end{aligned} \tag{3.5}$$

Then the coefficient of $G(u,z)$

$$\left[u^k z^n\right] G(u,z) = \left[{n \atop k}\right] \frac{2^k}{n!}. \tag{3.6}$$

Let us consider a ternary search tree. Ternary search trees have application in 3-valued logic, and is useful in deducing various parameters of *3-Huffman* trees [5]. Expected number of external nodes at level $k$ of a search tree having $n$ internal nodes satisfies the following recurrence relation [5].

$$E^{(3)}(k,n) = \frac{3}{n(n+1)/2} \sum_{i=0}^{n-1} (n-i)E^{(3)}(k-1,i). \tag{3.7}$$

The above recurrence relation corresponds to the generating function which satisfies the following differential equation:

$$z(1-z)^2 G_z''(u,z) + 2(1-z)^2 G_z'(u,z) - 6uG(u,z) = 0. \tag{3.8}$$

We could not solve equation (3.8) analytically. The following solution to (3.8) has been obtained using MAPLE.

$$\begin{aligned} G(u,z) &= 1 + 3uz + u(3u+2)z^2 \\ &+ \frac{1}{8}u\left(12u^2 + 32u + 12\right)z^3 \\ &+ \frac{1}{120}u\left(54u^3 + 360u^2 + 522u + 144\right)z^4 \\ &+ \cdots \end{aligned} \tag{3.9}$$

$G(u,0) = 1$ is the boundary condition to be satisfied by the solution.

Number of internal nodes, sum total of weights and expected external path length can be obtained by, respectively, replacing $u$ by 1, $u$ by $1/3$ and by replacing $u$ by 1 in $G'(u,z)$.

The above recurrence relation gives a lot more option than the recurrence relation that exists in the literature [3]. For example, the existing recurrences cannot be used to find the expected number of external nodes at a particular level of the tree. Parameters of ternary search trees can also be calculated from the relation.

## 4. SOME NEW FORMULAE

Stirling numbers, named after James Stirling, are close relatives of binomial coefficients. These numbers come in two flavours namely, Stirling number of the first and second kind. Stirling numbers of the first and second kinds combine themselves mutually in the following nice ways. Stirling number of the first kind counts the number of ways $n$ things can be arranged to have $k$ cycles, whereas Stirling number of the second kind counts the number of ways to partition a set of $n$ things into $k$ nonempty sets. Here again, we follow [3] for notations for Stirling numbers, where for Stirling number of the first kind square brackets are used, and for that of the second kind curly brackets are used. The formulae considered here are in the same spirit as in [3, p. 192].

FORMULA 1.

$$g(n) = \sum_k \left[{n \atop k}\right] f(k) \Leftrightarrow f(n) = \sum_k \left\{{n \atop k}\right\} (-1)^{n-k} g(k), \qquad \text{for } n > 0. \tag{3.10}$$

PROOF. Let us first prove that

$$g(n) = \sum_k \left[{n \atop k}\right] f(k) \Rightarrow f(n) = \sum_k \left\{{n \atop k}\right\} (-1)^{n-k} g(k).$$

Let $g(n) = \sum_k \left[{n \atop k}\right] f(k)$ be true, for all $n > 0$, then

$$\begin{aligned} \sum_k \left\{{n \atop k}\right\} (-1)^{n-k} g(k) &= \sum_k \left\{{n \atop k}\right\} (-1)^{n-k} \sum_j \left[{k \atop j}\right] f(j) \\ &= \sum_j f(j) \sum_k \left\{{n \atop k}\right\} \left[{k \atop j}\right] (-1)^{n-k} \\ &= \sum_j f(j)[n=j] \\ &= f(n). \end{aligned}$$

The proof for the reverse direction is similar since the relation between $f$ and $g$ is symmetric.

FORMULA 2.

$$g(n) = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} (-1)^{n-k} f(k) \Leftrightarrow f(n) = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} g(k), \qquad \text{for all } n > 0. \tag{3.11}$$

PROOF. Let us first prove that

$$g(n) = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} (-1)^{n-k} f(k) \Rightarrow f(n) = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} g(k).$$

Let $g(n) = \sum_k \left[ {n \atop k} \right] (-1)^{n-k} f(k)$ is true, for all $n > 0$, then

$$\begin{aligned}
\sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} g(k) &= \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} \sum_j \begin{bmatrix} k \\ j \end{bmatrix} (-1)^{k-j} f(j) \\
&= \sum_j f(j)(-1)^{n-j} \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{bmatrix} k \\ j \end{bmatrix} (-1)^{n-k} \\
&= \sum_j f(j)(-1)^{n-j} [n = j] \\
&= f(n).
\end{aligned}$$

The proof for the reverse direction is again similar for the same reason.

Now, let us see in what way they help us to have closed forms for other sums.
What is the value of $\sum_k \left\{ {n \atop k} \right\} (-1)^k$?
Now we focus to the formula one.

$$g(n) = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} f(k) \Leftrightarrow f(n) = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} g(k), \qquad \text{for } n > 0.$$

Here, if we assume $g(n) = n!$ and $f(n) = 1$, we reach

$$n! = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} \Leftrightarrow 1 = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^k k!.$$

So alternating sum of row of the Stirling number of first kind is 1.
We provide some similar identities where

$$\begin{aligned}
\begin{bmatrix} n+1 \\ m+1 \end{bmatrix} &= \sum_k \begin{bmatrix} n \\ k \end{bmatrix} \binom{k}{m} \Leftrightarrow \binom{n}{m} = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{bmatrix} k+1 \\ m+1 \end{bmatrix} (-1)^{n-k} \\
\begin{bmatrix} n-1 \\ m \end{bmatrix} &= \sum_k \begin{bmatrix} n \\ k \end{bmatrix} \binom{k-1}{m} (-1)^{m-1-k} \Leftrightarrow \binom{n-1}{m} (-1)^{m-n-1} \\
&= \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} \begin{bmatrix} k-1 \\ m \end{bmatrix} (-1)^{n-k},
\end{aligned}$$

$$\begin{aligned}
(n-1)^{n-1-m} [n-1 \geq m] &= \sum_k \begin{bmatrix} n \\ k \end{bmatrix} \begin{Bmatrix} k \\ m \end{Bmatrix} (-1)^{m-k} \Leftrightarrow \\
\begin{Bmatrix} n \\ m \end{Bmatrix} (-1)^{m-n} &= \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} (k-1)^{k-1-m} [k-1 \geq m],
\end{aligned}$$

$$z^{\overline{n}} = \sum_k \begin{bmatrix} n \\ k \end{bmatrix} z^k \Leftrightarrow z^n = \sum_k \begin{Bmatrix} n \\ k \end{Bmatrix} (-1)^{n-k} z^{\overline{k}}.$$

In [3], there is a closed form for the coefficients of $\left|{n \atop k}\right|$ in the representation of rising powers by falling powers.

$$x^{\overline{n}} = \sum_k \left|{n \atop k}\right| x^{\underline{k}}, \qquad \text{int}\, n \geq 0. \tag{3.12}$$

This results in a recurrence relation for $\left|{n \atop k}\right|$.

Here we would like to find a closed form for the coefficient of $\left\langle {n \atop k} \right\rangle$, known as Eulerian number, the number of permutations of $\{1, 2, \ldots, n\}$ that have $k$ ascents (see [3, p. 267]), in the representation of falling powers by the rising powers:

$$x^{\underline{n}} = \sum_k \left\langle {n \atop k} \right\rangle x^{\overline{k}}, \qquad \text{int}\, n \geq 0, \tag{3.13}$$

$$\begin{aligned} x^{\underline{n}} &= (x-n+1)^{\overline{n}} \\ &= \sum_k \binom{n}{k} x^{\overline{k}} (-n+1)^{\overline{n-k}}. \end{aligned} \tag{3.14}$$

Equating coefficient of (3.13) and (3.14), we get

$$\left\langle {n \atop k} \right\rangle = \binom{n}{k} (-n+1)^{\overline{n-k}}. \tag{3.15}$$

Now, let us find a recurrence relation for $\left\langle {n \atop k} \right\rangle$

$$\begin{aligned} \left\langle {n \atop k} \right\rangle &= \binom{n}{k} (-n+1)^{\overline{n-k}} \\ &= \binom{n}{k} (-n+2)^{\overline{n-k}} \frac{1-n}{1-k} \qquad [k \neq 1] \\ &= \left\{ \binom{n-1}{k-1} (-n+2)^{\overline{n-k}} + \binom{n-1}{k} (-n+2)^{\overline{n-k}} \right\} \frac{1-n}{1-k} \qquad [k \neq 0] \\ &= \left\{ \left\langle {n-1 \atop k-1} \right\rangle + \binom{n-1}{k} (-n+2)^{\overline{n-k-1}} (1-k) \right\} \frac{1-n}{1-k} \qquad [k \neq 0] \\ &= \left\{ \left\langle {n-1 \atop k-1} \right\rangle + \left\langle {n-1 \atop k} \right\rangle (1-k) \right\} \frac{1-n}{1-k} \qquad [k \neq 0] \\ &= (1-n-k) \left\langle {n-1 \atop k} \right\rangle + k \left\langle {n-1 \atop k} \right\rangle + \frac{1-n}{1-k} \left\langle {n-1 \atop k-1} \right\rangle \qquad [k \neq 1] \\ &= (1-n-k) \left\langle {n-1 \atop k} \right\rangle + k \binom{n-1}{k} (-n+2)^{\overline{n-k-1}} + \frac{1-n}{1-k} \left\langle {n-1 \atop k-1} \right\rangle \qquad [k \neq 1] \\ &= (1-n-k) \left\langle {n-1 \atop k} \right\rangle + k \frac{n-k}{k(1-k)} \binom{n-1}{k-1} (-n+2)^{\overline{n-k}} \\ &\quad + \frac{1-n}{1-k} \left\langle {n-1 \atop k-1} \right\rangle \qquad [k \neq 1] \\ &= (1-n-k) \left\langle {n-1 \atop k} \right\rangle + \frac{n-k}{1-k} \left\langle {n-1 \atop k-1} \right\rangle + \frac{1-n}{1-k} \left\langle {n-1 \atop k-1} \right\rangle \qquad [k \neq 1] \end{aligned} \tag{3.16}$$

$$\left\langle {n \atop k} \right\rangle = (1-n-k) \left\langle {n-1 \atop k} \right\rangle + \left\langle {n-1 \atop k-1} \right\rangle, \qquad [\text{for all integer } n, k]$$

$$f(n) = \sum_k \left|{n \atop k}\right| g(k) \Leftrightarrow g(n) = \sum_k \left\langle {n \atop k} \right\rangle f(k) \qquad \text{int}\, n \geq 0.$$

Let assume the left side is true, then

$$
\begin{aligned}
\sum_k \left\langle {n \atop k} \right\rangle f(k) &= \sum_k \left\langle {n \atop k} \right\rangle \sum_j \left| {k \atop j} \right| g(j) \\
&= \sum_j g(j) \sum_k \left\langle {n \atop k} \right\rangle \left| {k \atop j} \right| \\
&= \sum_j g(j) \sum_k \binom{n}{k} (-n+1)^{\overline{n-k}} \binom{k}{j} (k-1)^{\underline{k-j}} \\
&= \sum_j g(j) \sum_k \binom{n}{k} \binom{k}{j} (-1)^{n-k} (n-1)^{\underline{n-k}} (k-1)^{\underline{k-j}} \\
&= \sum_j g(j) \sum_k \binom{n}{k} \binom{k}{j} (-1)^{n-k} (n-1)^{\underline{n-j}} \\
&= \sum_j g(j) (-1)^n (n-1)^{\underline{n-j}} \sum_k \binom{n}{k} \binom{k}{j} (-1)^k \\
&= \sum_j g(j) (-1)^n (n-1)^{\underline{n-j}} (-1)^n \binom{0}{j-n} \\
&= g(n) \\
\left| {n \atop k} \right| &= \binom{n}{k} (n-1)^{\underline{n-k}} \\
\left\langle {n \atop k} \right\rangle &= \binom{n}{k} (1-n)^{\overline{n-k}} \\
&= \binom{n}{k} (-1)^{n-k} (n-1)^{\underline{n-k}}.
\end{aligned}
$$

Hence,

$$
\left\langle {n \atop k} \right\rangle = (-1)^{n-k} \left| {n \atop k} \right| .
$$

As [3] provides an inversion formula for $|\;|$ as

$$
\left| {n \atop k} \right| = \left| {-k \atop -n} \right| ,
$$

so do we for $\langle\;\rangle$ as

$$
\begin{aligned}
\left\langle {n \atop k} \right\rangle &= (-1)^{n-k} \left| {n \atop k} \right| = (-1)^{-k-(-n)} \left| {-k \atop -n} \right| , \\
\left\langle {n \atop k} \right\rangle &= \left\langle {-k \atop -n} \right\rangle .
\end{aligned}
\tag{3.17}
$$

FORMULA 3. How many ways can $n$ people be grouped where each group contains exactly $k$ people?

$$
{}^kP_n = \begin{cases}
1, & \text{if } k = n, \\
0, & \text{if } k > n, \\
\dbinom{n-1}{k-1} {}^kP_{n-k}, & \text{if } k \text{ divides } p, \\
\dbinom{n}{\left\lfloor \frac{n}{k} \right\rfloor * k} {}^kP_{\lfloor n/k \rfloor * k}, & \text{if } k \text{ does not divide } p.
\end{cases}
$$

## 5. CONCLUSION

In this paper, we have applied Stirling numbers to express values of some parameters of binary search trees that are not available in the literature so far as our knowledge goes. We have also derived some formulae relating Stirling numbers of the first and second kinds.

## REFERENCES


1. D.E. Knuth, *The Art of Computer Programming*, Vol. III: Sorting and Searching, Addison-Wesley, Reading, MA, (1973).
2. M.R. Brown, A partial analysis of random height-balanced trees, *SIAM J. on Computing* **8** (1), 33–41, (Feb. 1979).
3. R.L. Graham, D.E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley Publishing, (1994).
4. T.M. Islam, R.A. Chowdhury, S.K. Nath and M. Kaykobad, On parameters of search trees, In *Proceedings of International Conference on Computer and Information Technology*, pp. 50–53, (1999).
5. A.H. Manob and M. Kaykobad, Stirling numbers and two inversion formulas, In *International Conference on Computer and Information Technology*, Held at Dhaka University during December 28–29, pp. 6–8, (2001).
6. G.M. Adel'son-Vel'skii and E.M. Landis, An algorithm for the organization of information, *Cdokladi Akademia Nauk SSSR* **146** (2), 263–266, (1962); English translation, *Soviet Math. Doklady* **3**, 1259–1263, (1962).
7. M.E. Reingold and W.J. Hansen, *Data Structures*, 1st Indian Edition, CBS Publishers and Distributors, (1988).
8. A.C.C. Yao, On random 2-3 trees, *Acta Informatica* **9** (2), 159–170, (1978).